1/D

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/684,975	10/14/2003	David Alford Burgoon	BAT 0072 PA/40078.248	9946

7590 06/28/2006

DINSMORE & SHOHL LLP
One Dayton Center
Suite 500
Dayton, OH 45402-2023

EXAMINER	
GORTAYO, DANGELINO N	

ART UNIT	PAPER NUMBER
2168	

DATE MAILED: 06/28/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/684,975	BURGOON ET AL.	
	Examiner	Art Unit	
	Dangelino N. Gortayo	2168	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *14 October 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-111* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-96 and 111* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☒ Claim(s) *1-111* are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *14 October 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *1/16/04; 7/12/04*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-111 are pending.

2. Initialed and dated copies of Applicant's IDS forms 1449, filed 1/16/2004 and 7/12/2004, are attached to the instant Office action.

Election/Restrictions

3. Restriction to one of the following inventions is required under 35 U.S.C. 121:

   I. Claims 1-96, and 111, drawn to a process to create an information reservoir composed of table from data sources, classified in class 707/100.

   II. Claims 97-110 drawn to a method to translate queries directed at data sources into queries for a data source representation, classified in class 707, subclass 4.

4. The inventions are distinct, each from the other because of the following reasons:

Inventions Group I and II are related as combination and subcombination. Inventions in this relationship are distinct if it can be shown that (1) the combination as claimed does not require the particulars of the subcombination as claimed for patentability, and (2) that the subcombination has utility by itself or in other combinations (MPEP § 806.05(c)). In the instant case, the combination as claimed does not require the particulars of the subcombination as claimed because the creation

of an information reservoir database representation does not require the translation of queries to be used. The subcombination has separate utility such as translating queries from various data sources into a uniform representation for better communication.

5. Because these inventions are independent or distinct for the reasons given above and have acquired a separate status in the art in view of their different classification, restriction for examination purposes as indicated is proper.

6. Because these inventions are independent or distinct for the reasons given above and search required for Group I is not required for Group II, restriction for examination purposes as indicated is proper.

7. A telephone call was made to Attorney James E. Bayer (Registration Number 39,564) on 5/12/2006 to request an oral election to the above restriction requirement, group I is elected with traverse.

Claim Rejections - 35 USC § 102

8. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

9. Claims 1-52, 57-96, and 111 are rejected under 35 U.S.C. 102(e) as being anticipated by Chaudhuri et al. ("Chaudhuri" US Patent 6,532,458 B1 filed 3/15/1999)

As per claim 1, Chaudhuri teaches "A computer-implemented information reservoir creation process wherein:" (see Abstract)

"a table collection is constructed from a data source;" (column 7 lines 7-20, wherein a sampling of records from a database is taken)

"said table collection includes a subset of tables designated as sampling initiation tables;" (column 10 lines 42-57)

"each table in said table collection is a member of either a directly-sampled table set or a descendent-sampled table set;" (column 11 lines 52-67, wherein the tables are from tuples directly or stemming from tuples that are obtained)

"said directly-sampled table set is characterized by tables that are either sampling initiation tables or ancestor tables to one or more sampling initiation tables;" (column 11 lines 20-27)

"said descendant-sampled table set is characterized by tables that are descendant tables to a sampling initiation table;" (column 11 lines 32-39)

"said table collection is characterized by a table collection schema equivalent to a data source schema of said data source, with the exception that a list of attributes for each table of said directly-sampled table set includes an additional attribute containing actual rate of inclusion values;" (column 12 lines 13-19, wherein samples are based on tuple data and a probability of sampling based on a weight)

"each tuple included in said table collection is equivalent to one and only one tuple in the corresponding table of said data source;" (column 12 lines 1-7, wherein the each tuple of the relation is evaluated)

"an actual rate of inclusion value stored with a select data source tuple and included in a directly-sampled table of said table collection represents the probability that a randomly selected table collection produced by the process will contain said select data source tuple." (column 12 lines 20-27, wherein a weight of a relation determines sampling of tuples)

As per claim 2, Chaudhuri teaches "each tuple included in said table collection is equivalent to one and only one tuple in the corresponding table of said data source." (column 12 lines 1-7, wherein the each tuple of the relation is evaluated)

As per claim 3, Chaudhuri teaches "each tuple included in said table collection is equivalent to one and only one tuple in the corresponding table of said data source after elimination of said actual rate of inclusion value." (column 11 lines 62-67, wherein the tuples in the sample correspond to the database tuples)

As per claim 4, Chaudhuri teaches "said table collection includes all ancestor tuples of each tuple included in any directly-sampled table of the table collection." (column 16 lines 6-13)

As per claim 5, Chaudhuri teaches "said table collection includes all descendant tuples of each tuple included in any sampling initiation table of the table collection." (column 16 lines 6-13)

As per claim 6, Chaudhuri teaches "said probability that a randomly selected table collection produced by the process will contain a given data source tuple in a descendant-sampled table is equal to the actual rate of inclusion stored with a corresponding single ancestor tuple residing in a sampling initiation table." (column 5 line 62 – column 6 line 4)

As per claim 7, Chaudhuri teaches "no pair of data source tuples within any select tuple set taken from directly-sampled tables has an ancestor-descendant relationship." (column 16 line 18-21)

As per claim 8, Chaudhuri teaches "the probability that a randomly selected table collection produced by the process will contain all of the tuples in said select tuple set is equal to the product of the corresponding actual rates of inclusion associated with each of the individual data source tuples." (column 5 line 62 – column 6 line 4)

As per claim 9, Chaudhuri teaches "A computer-implemented method for constructing a representation from a data source in order to provide relatively quick response to queries related to information in said data source, wherein said data source has a plurality of tuples stored in said data source and a data source schema that includes defined relationships among at least a subset of the tuples in the data source, said method comprising:" (see Abstract)

"creating said representation by copying at least a subset of said data source schema to define a representation schema;" (column 12 lines 13-19, wherein samples are based on tuple data and a probability of sampling based on a weight)

"adding additional data to said representation that represents information that is not in said data source;" (column 12 lines 13-19, wherein the weights are not present in the data source and are attached in the sampling process)

"defining tuples of interest within said data source and a degree of interest for each tuple of interest;" (column 12 lines 20-27, wherein a weight of a relation determines sampling of tuples)

"sampling tuples from said tuples of interest into said representation based upon said degree of interest in a manner that preserves at least a subset of said relationships among tuples in the data source;" (column 12 lines 40-58)

"and storing values in the representation that relate to a likelihood that each tuple sampled into said representation would be sampled into the representation if the sampling process were to be repeated. (column 13 lines 19-30, wherein the reservoir array shows tuple order and is collected through sampling)

As per claim 10, Chaudhuri teaches "said data source is a table collection." (column 7 lines 7-20, wherein a sampling of records from a database is taken)

As per claim 11, Chaudhuri teaches "said table collection is a relational database and said defined relationships among tuples are foreign key relationships." (column 1 lines 36-45)

As per claim 12, Chaudhuri teaches "said representation schema comprises a logically limited subset of said data source schema." (column 10 lines 22-33, wherein the sampling is from a single pass of a database)

As per claim 13, Chaudhuri teaches "said additional data for an individual tuple includes selected aggregates of descendant tuples." (column 10 lines 58-64, wherein the individual tuple is read based on its relations)

As per claim 14, Chaudhuri teaches "said representation is to be used to respond to queries against a parent table that are restricted to parents of a particular kind of child type; and said representation further includes data added to said representation that is indicative of whether a select tuple in said parent table is associated with said particular kind of child type." (column 9 lines 24-32)

As per claim 15, Chaudhuri teaches "said tuples of interest are defined by a plurality of attributes and only a subset of said plurality of attributes are copied for each tuple into said representation." (column 17 lines 50-58)

As per claim 16, Chaudhuri teaches "said tuples of interest are defined by associating with each tuple of interest a target rate of inclusion greater than zero and said degree of interest is indicated by the magnitude of the target rate of inclusion." (column 13 lines 31-47, wherein the weight is greater than zero and the sum of the weights increases with the weight)

As per claim 17, Chaudhuri teaches "determining said target rate of inclusion comprises taking a minimum of the quantity one and the result of dividing the number of tuples desired in the representation by the total number of tuples in the data source that are to be considered for sampling." (column 12 lines 28-36, wherein the variable p is a number representing the number of tuples yet to be sampled)

As per claim 18, Chaudhuri teaches "said representation is biased by assigning a higher target rate of inclusion for a subset of said tuples of interest." (column 12 lines 12-19)

As per claim 19, Chaudhuri teaches "determining said target rate of inclusion comprises taking the minimum of the quantity one and the result of dividing the number of tuples desired in the representation by a number of subpopulations, and dividing that result by the number of tuples in each subpopulation." (column 14 lines 21-38)

As per claim 20, Chaudhuri teaches "identifying one or more real-valued attributes of interest in said data source;" (column 21 lines 37-43, "joint attribute values") "clustering said data source based upon said real-valued attributes of interest;" (column 21 lines 47-52) "and partitioning said population into subpopulations based upon said clustering, wherein said rates of inclusion are assigned to tuples by subpopulation." (column 21 line 65 – column 22 line 9)

As per claim 21, Chaudhuri teaches "said target rate of inclusion is set to its maximum value for tuples containing attribute values that have a high degree of influence on anticipated query results." (column 14 lines 56-64)

As per claim 22, Chaudhuri teaches "knowledge of an anticipated workload is encoded into a first set of queries that are representative of said knowledge of said anticipated workload to derive weighting factors used to establish said target rates of inclusion." (column 9 lines 33-46 and column 10 lines 35-41, wherein a SAMPLE query is entered including weighting factor)

As per claim 23, Chaudhuri teaches "determining a training set of queries defining a reservoir training set;" (column 10 line 35-41, relation R of tuples) "associating a set of aggregates with each training query;" (column 12 lines 13-19) "collecting said aggregates into a superset;" (column 12 lines 13-19) "determining weights for said aggregates in said superset to reflect the importance to users of said representation;" (column 12 lines 20-27) "determining a tuning parameter from said weights;" (column 12 lines 28-36) "partitioning a sampling population into at least those tuples in the scope of said aggregates, and those tuples outside the scope of said aggregates;" (column 15 lines 39-47) and determining target rates of inclusion for the tuples in each group." (column 15 lines 48-55)

As per claim 24, Chaudhuri teaches "said target rates of inclusion for said tuples in the scope of said aggregates in said superset are chosen to minimize the variances of aggregate estimates computed from the representation." (column 15 line 56-65)

As per claim 25, Chaudhuri teaches "said rate of inclusion for tuples participating in sums has the property that tuples with attribute values that are relatively large in magnitude are assigned larger target rates of inclusion." (column 16 line 65 – column 16 line 3)

As per claim 26, Chaudhuri teaches "said rate of inclusion for tuples participating in averages has the property that tuples with outlying attribute values are assigned larger target rates of inclusion." (column 16 lines 22-29)

As per claim 27, Chaudhuri teaches "controlling the size of said representation" (column 16 lines 22-29)

As per claim 28, Chaudhuri teaches "said tuple preference factor is selected between the values of zero and the quotient defined by the number of said tuples of interest in said data source divided by said target number of tuples such that the sum of all tuple preference factors equals the number of said tuples of interest." (column 14 lines 50-53)

As per claim 29, Chaudhuri teaches "said target rate of inclusion for a select tuple among said tuples of interest is computed by multiplying said target number of tuples by said tuple preference factor, and dividing that product by the number of said tuple of interest." (column 14 lines 25-38)

As per claim 30, Chaudhuri teaches "the space required by said representation is determined comprising: determining an average tuple inclusion probability;" (column 17 line 64 – column 18 line 4) "and approximating said space by multiplying said average tuple inclusion probability by the sum of a first space required to store the actual tuples in said data source to be considered for sampling and a second space required to store auxiliary structures whose sizes are proportional to said first space, and adding to that product, a third space required to store auxiliary structures whose sizes are not proportional to said first space." (column 18 lines 5-15)

As per claim 31, Chaudhuri teaches "said average tuple inclusion probability is determined by dividing a target number of tuples in said representation by the number of said tuples of interest in said data source." (column 21 lines 22-37)

As per claim 32, Chaudhuri teaches "determining an estimate of the size of said representation" (column 22 lines 19-27)

As per claim 33, Chaudhuri teaches "the number of child tuples is obtained using a frequency table." (column 22 lines 53-58)

As per claim 34, Chaudhuri teaches "the number of child tuples is obtained using an index on the foreign key linking said relationship to said child tuples." (column 22 lines 53-58)

As per claim 35, Chaudhuri teaches "said average actual inclusion probability of said parent table is calculated as a weighted average of the average inclusion probability of each subset of parent tuples having the same number of child tuples." (column 22 lines 61-67)

As per claim 36, Chaudhuri teaches "ancestor tuples, both within and outside of said tuples of interest, of at least a subset of tuples selected into said representation may be given a higher chance of being selected into said representation." (column 23 lines 35-42)

As per claim 37, Chaudhuri teaches "ancestor tuples of at least a subset of tuples selected into said representation are necessarily included in said representation." (column 24 lines 1-6)

As per claim 38, Chaudhuri teaches "descendant tuples, both within said tuples of interest and outside of said tuples of interest, of at least a subset of tuples selected into said representation are given a higher chance of being selected into said representation." (column 23 lines 35-42)

As per claim 39, Chaudhuri teaches "descendant tuples of at least a subset of tuples selected into said representation are included in said representation." (column 24 lines 1-6)

As per claim 40, Chaudhuri teaches "an adjusted rate of inclusion is determined for each tuple of interest, said adjusted rate comprising possible contributions from said degree of interest in said tuple, from the results of sampling ancestor tuples of said tuple, and from the results of sampling descendant tuples of said tuple, and the act of sampling an individual tuple among said tuples of interest comprises: considering a select tuple from said tuples of interest;" (column 20 lines 36-46) "simulating a trial in which an event occurs with probability equal to the adjusted rate of inclusion;" (column 21 lines 1-7) "determining whether or not the event has occurred;" (column 21 lines 1-7) "and copying select tuple into said representation if and only if said event occurs." (column 24 lines 44-51)

As per claim 41, Chaudhuri teaches "said event is that a uniform random number on the open interval (0,1) is less than said adjusted rate of inclusion." (column 15 lines 32-35)

As per claim 42, Chaudhuri teaches "said trials for any pair of tuples within a table are simulated independently." (column 24 lines 11-23)

As per claim 43, Chaudhuri teaches "said act of determining an adjusted rate of inclusion comprises: assigning a target rate of inclusion to the select tuple of interest;" (column 21 lines 52-59) "computing an induced rate of inclusion that represents the rate of inclusion induced by any descendant or ancestor tuples of said select tuple, said

induced rate of inclusion set to zero if said select tuple has no descendants or ancestors;" (column 22 lines 10-18) "and computing an adjusted rate of inclusion based upon said target rate of inclusion and said induced rate of inclusion, wherein said tuples of interest are sampled based upon said adjusted rate of inclusion." (column 22 lines 41-52)

As per claim 44, Chaudhuri teaches "said induced rate of inclusion and said adjusted rate of inclusion are computed only if said select tuple is related to any descendant or ancestor tuples." (column 22 lines 48-52)

As per claim 45, Chaudhuri teaches "said tuple of interest is associated with descendant and ancestor tuples that are partitioned into subgroups and said induced rate of inclusion is determined by: computing an induced rate of inclusion for each subgroup based on the actual rates of inclusion associated with descendant and ancestor tuples in the subgroup;" (column 22 lines 53-58) "and computing an overall induced rate of inclusion from the component rates of inclusion induced by each subgroup." (column 22 lines 56-61)

As per claim 46, Chaudhuri teaches "said data source is dynamic with new tuples arriving over time, wherein each subgroup comprises sibling tuples partitioned by their arrival time into said data source." (column 13 lines 11-18, wherein sampling over one pass is for changes in the database)

As per claim 47, Chaudhuri teaches "said data source is distributed over a number of computer devices greater than one, wherein each subgroup comprises sibling tuples partitioned by computer devices." (column 9 lines 1-24)

As per claim 48, Chaudhuri teaches "said adjusted rate of inclusion is equal to the greater of zero and the result of the induced rate of inclusion subtracted from the target rate of inclusion divided by the result of subtracting the induced rate of inclusion from one." (column 22 lines 4-9)

As per claim 49, Chaudhuri teaches "said select tuple is sampled at the time said select tuple's corresponding table is sampled at a sampling rate equal to the adjusted rate of inclusion." (column 22 lines 10-18)

As per claim 50, Chaudhuri teaches "said select tuple is not sampled if said induced rate of inclusion is greater than or equal to said target rate of inclusion." (column 22 lines 41-48)

As per claim 51, Chaudhuri teaches "an actual rate of inclusion is computed for each tuple selected into said representation, said actual rate of inclusion reflecting all opportunities for said tuple to be included in said representation." (column 12 lines 40-58)

As per claim 52, Chaudhuri teaches "said actual rate of inclusion is part of said additional data added to said representation." (column 12 lines 13-19, wherein the weights are attached in the sampling process)

As per claim 57, Chaudhuri teaches "said representation defines a second representation that is a subsample of a first representation" (column 16 lines 30-33)

As per claim 58, Chaudhuri teaches "said representation defines a third representation that is the union of a first representation and a second representation" (column 22 lines 28-37)

As per claim 59, Chaudhuri teaches "said representation defines a third representation that is the intersection of a first representation and a second representation" (column 22 lines 28-37)

As per claim 60, Chaudhuri teaches "said representation defines a first representation and said method further comprises establishing a maximum size for said representation and when said maximum size is exceeded, reducing the size of said representation" (column 17 lines 34-40)

As per claim 61, Chaudhuri teaches "said subsample target rate of inclusion is equal to the desired size of said second representation divide by the size of said first representation." (column 17 lines 50-58)

As per claim 62, Chaudhuri teaches "said size is measured in units of numbers of tuples." (column 18 lines 5-15)

As per claim 63, Chaudhuri teaches "said size is measured in terms of bytes of disk storage space." (column 21 lines 8-17, wherein size of the sample is in partitions with an index, residing in memory)

As per claim 64, Chaudhuri teaches updating said representation in view of a change occurring to said data source (column 10 lines 22-33, wherein the sampling is from a single pass of a database)

As per claim 65, Chaudhuri teaches "changes are identified based upon a batch driven process." (column 10 lines 22-27)

As per claim 66, Chaudhuri teaches "changes are identified in at least near real time." (column 10 lines 22-33)

As per claim 67, Chaudhuri teaches updating said representation in view of added tuples occurring to said data source (column 10 lines 22-33, wherein the sampling is from a single pass of a database)

As per claim 68, Chaudhuri teaches "adjusting select inclusion probabilities over time in response to modifications to said data source." (column 12 lines 13-19)

As per claim 69, Chaudhuri teaches "constructing a buffer that substantially mirrors said representation schema;" (column 24 lines 23-29) "copying said added tuples into said buffer;" (column 11 lines 47-51) "copying any ancestor tuples and descendant tuples related to each added tuple into said buffer;" (column 11 lines 47-51) "assigning a rate of inclusion to said added tuples in said buffer;" (column 12 lines 20-27) "and sampling tuples from said buffer into said representation based upon associated rates of inclusion." (column 24 lines 32-35)

As per claim 70, Chaudhuri teaches maintaining the relative size of said representation" (column 10 lines 3-12)

As per claim 71, Chaudhuri teaches " said representation is incrementally updated as said data source is updated." (column 10 lines 22-33)

As per claim 72, Chaudhuri teaches "said representation is continually rebuilt." (column 10 lines 22-33)

As per claim 73, Chaudhuri teaches "said representation is continually rebuilt by defining logical partitions of tables of said data source, ordering said logical partitions, and, for each logical partition: loading a select partition into a buffer; adding tuples to said buffer as necessary for said buffer to contain the closure of said select partition;

sampling said buffer; joining the sampled buffer with said representation; and updating rates of inclusion of tuples sampled from said buffer." (column 21 lines 22-37)

As per claim 74, Chaudhuri teaches "said representation is subsampled to control the size of the rebuilt representation." (column 22 lines 19-27)

As per claim 75, Chaudhuri teaches "answering queries against said data source with approximate answers computed from said representation." (column 24 lines 1-6)

As per claim 76, Chaudhuri teaches "providing a variance with said approximate answer." (column 24 lines 1-6)

As per claim 77, Chaudhuri teaches "providing a confidence interval for the exact answer with said approximate answer." (column 19 lines 24-30)

As per claim 78, Chaudhuri teaches "A system for constructing a representation from a data source in order to provide response to queries related to information in said data source, wherein said data source has a plurality of tuples stored in said data source and a data source schema that includes defined relationships among at least a subset of the tuples in the data source, said system comprising:" (see Abstract)

"at least one processor;" (column 8 lines 29-37)

"at least one storage device communicably coupled to said at least one processor arranged to store said data source and said representation;" (column 8 lines 29-37)

"and software executable by said at least one processor for: creating said representation by copying at least a subset of said data source schema to define a representation schema;" (column 8 lines 11-15)

"adding additional data to said representation that represents information that is not in said data source;" (column 12 lines 13-19, wherein the weights are not present in the data source and are attached in the sampling process)

"defining tuples of interest within said data source and a degree of interest for each tuple of interest;" (column 12 lines 20-27, wherein a weight of a relation determines sampling of tuples)

"sampling tuples from said tuples of interest into said representation based upon said degree of interest in a manner that preserves at least a subset of said relationships among tuples in the data source;" (column 12 lines 40-58)

"and storing values in the representation that relate to the likelihood that each tuple sampled into said representation would be sampled into the representation if the sampling process were to be repeated." (column 13 lines 19-30, wherein the reservoir array shows tuple order and is collected through sampling)

As per claim 79, Chaudhuri teaches "said software implements a designer component for: interacting with a user; and defining parameters used to construct said representation based upon said parameters." (column 24 lines 1-6 and column 9 lines 47-55)

As per claim 80, Chaudhuri teaches "said designer component provides a user with a list of distinct valid values of categorical attributes from dimension defining tables

and/or a list of valid value ranges for real-valued attributes; and those subsets of tuples in said data source not associated with categorical values or value ranges that are selected by the user are marked for exclusion from said representation." (column 8 lines 64-67)

As per claim 81, Chaudhuri teaches " said software implements a designer component for: interacting with a user; and defining parameters used to construct a collection of scaled representations based upon said parameters." (column 24 lines 1-6 and column 9 lines 47-55)

As per claim 82, Chaudhuri teaches "said software is configured to construct a collection of scaled representations by first constructing a largest representation and then subsampling said largest representation." (column 10 lines 22-33)

As per claim 83, Chaudhuri teaches "said software implements a designer component for interacting with a user to allow said user to adjust the balance of tuples in said representation and said software constructs said representation based upon said adjustment." (column 24 lines 1-6 and column 9 lines 47-55)

As per claim 84, Chaudhuri teaches "said software implements an analyst component for: intercepting an original query;" (column 9 lines 25-32) "remapping said original query into a format compatible with said representation;" (column 9 lines 33-46) "applying said remapped query against said representation;" (column 9 lines 33-46) "and providing the results of the remapped query in response to said original query." (column 9 lines 42-46)

As per claim 85, Chaudhuri teaches "said results of the remapped query include one or more approximate answers." (column 24 lines 1-6)

As per claim 86, Chaudhuri teaches "said results of the remapped query include a variance with each approximate answer." (column 24 lines 1-6)

As per claim 87, Chaudhuri teaches "said results of the remapped query include a confidence interval for the exact answer with each approximate answer." (column 19 lines 24-30)

As per claim 88, Chaudhuri teaches "said software implements a builder component for constructing multiple representations of said data source and said analyst component is further configured for selecting between said multiple representations to select an optimal representation from said multiple representations to apply said remapped query against." (column 19 lines 35-44)

As per claim 89, Chaudhuri teaches "said software is further configured to construct multiple scaled versions of said representation and said software is further capable of applying said remapped query against a select one of said multiple scaled versions of said representation." (column 16 lines 30-33)

As per claim 90, Chaudhuri teaches "said multiple representations constructable by said builder component are selected from the group consisting of sampling, pre-computed aggregates, histograms, wavelets, data cubes, and data clouds." (column 9 lines 17-24)

As per claim 91, Chaudhuri teaches "said software implements a reporter component for outputting one or more approximate answers to said original query." (column 24 lines 1-6)

As per claim 92, Chaudhuri teaches "said reporter component optionally outputs a variance with each approximate answer." (column 24 lines 1-6)

As per claim 93, Chaudhuri teaches "said variance is provided by the reporter component as hidden metadata." (column 24 lines 1-6)

As per claim 94, Chaudhuri teaches "said reporter component optionally outputs a confidence interval for the exact answer with each approximate answer." (column 19 lines 24-30)

As per claim 95, Chaudhuri teaches " said confidence interval is provided by the reporter component as hidden metadata." (column 19 lines 24-30)

As per claim 96, Chaudhuri teaches "A computer readable medium including program code representing computer implemented operations for constructing a representation from a data source in order to provide relatively quick response to queries related to information in said data source, wherein said data source has a plurality of tuples stored in said data source and a data source schema that includes defined relationships among at least a subset of the tuples in the data source, said operations comprising:" (see Abstract)

"creating said representation by copying at least a subset of said data source schema to define a representation schema;" (column 12 lines 13-19, wherein samples are based on tuple data and a probability of sampling based on a weight)

"adding additional data to said representation that represents information that is not in said data source;" (column 12 lines 13-19, wherein the weights are not present in the data source and are attached in the sampling process)

"defining tuples of interest within said data source and a degree of interest for each tuple of interest;" (column 12 lines 20-27, wherein a weight of a relation determines sampling of tuples)

"sampling tuples from said tuples of interest into said representation based upon said degree of interest in a manner that preserves at least a subset of said relationships among tuples in the data source;" (column 12 lines 40-58)

"and storing values in the representation that relate to a likelihood that each tuple sampled into said representation would be sampled into the representation if the sampling process were to be repeated. (column 13 lines 19-30, wherein the reservoir array shows tuple order and is collected through sampling)

As per claim 111, Chaudhuri teaches "A computer readable medium including program code representing computer implemented operations for constructing a representation from a data source in order to provide relatively quick response to queries related to information in said data source, wherein said data source has a plurality of tuples stored in said data source and a data source schema that includes

defined relationships among at least a subset of the tuples in the data source, said operations comprising:" (see Abstract)

"creating said representation by copying at least a subset of said data source schema to define a representation schema;" (column 12 lines 13-19, wherein samples are based on tuple data and a probability of sampling based on a weight)

"adding additional data to said representation that represents information that is not in said data source;" (column 12 lines 13-19, wherein the weights are not present in the data source and are attached in the sampling process)

"defining tuples of interest within said data source and a degree of interest for each tuple of interest;" (column 12 lines 20-27, wherein a weight of a relation determines sampling of tuples)

"sampling tuples from said tuples of interest into said representation based upon said degree of interest in a manner that preserves at least a subset of said relationships among tuples in the data source;" (column 12 lines 40-58)

"and storing values in the representation that relate to a likelihood that each tuple sampled into said representation would be sampled into the representation if the sampling process were to be repeated. (column 13 lines 19-30, wherein the reservoir array shows tuple order and is collected through sampling)

Claim Rejections - 35 USC § 103

10. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

11. Claims 53-55 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chaudhuri et al. ("Chaudhuri" US Patent 6,532,458 B1 filed 3/15/1999) in view of Acharya et al. ("Acharya" US Patent 6,912,524 B2 filed 8/12/2002, a divisional of Application No. 09/480,261 filed 1/11/2000)

As per claim 53, Chaudhuri teaches "copying each tuple selected through sampling into said representation;" (column 24 lines 44-51) "and optionally copying ancestor and descendant tuples associated with each tuple selected through sampling into said representation." (column 16 lines 6-13). Chaudhuri does not teach "said method further comprises: representing said subset of said data source schema as a directed, acyclic graph having tables as vertices and table relationships as directed edges, said edges defining ancestor-descendant relationships between tuples in said data source; traversing said vertices of said acyclic graph; sampling each tuple associated with said vertices as each vertex is visited;"

Acharya teaches "said method further comprises: representing said subset of said data source schema as a directed, acyclic graph having tables as vertices and table relationships as directed edges, said edges defining ancestor-descendant relationships between tuples in said data source; traversing said vertices of said acyclic

graph; sampling each tuple associated with said vertices as each vertex is visited;" (Figure 2 and column 10 lines 25-44). It would have been obvious at the time of the invention for one of ordinary skill in the art to combine Chaudhuri's method of constructing a sampling representation from a data source with Acharya's ability to represent sampling data in a directed acyclic graph. This gives the user the advantage of representing a sampling representation in another way, extending the features of Chaudhuri's patents. The motivation for doing so would be produce approximate answers with a quicker response time based on a sampling representation. (column 1 lines 47-50)

As per claim 54, Chaudhuri teaches "wherein said data source is a table collection." (column 7 lines 7-20, wherein a sampling of records from a database is taken)

As per claim 55, Chaudhuri teaches "said table collection is a relational database and said ancestor-descendant relationships between tuples are foreign key relationships." (column 1 lines 36-45)

As per claim 56, Chaudhuri and Acharya are disclosed as per claim 53 above. Additionally, Acharya teaches "said act of traversing said vertices comprises: identifying a subset of the vertices as sampling initiation points; performing a breadth-first traversal of those vertices identified as sampling initiation points; traversing all vertices that can be reached from a sampling initiation point via pathways that follow the direction of said directed edges; and traversing all vertices that can be reached from a sampling initiation

point via pathways that follow the opposite direction of said directed edges." (column 10 line 53 – column 11 line 2)

Conclusion

12. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Castelli et al. (US Patent 6,122,628)

Ayukawa et al. (US Patent 6,510,457 B1)

Fayyad et al. (6,633,882 B1)

13. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dangelino N. Gortayo whose telephone number is (571)272-7204. The examiner can normally be reached on M-F 7:30-4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Tim T. Vo can be reached on (571)272-3642. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Dangelino N. Gortayo
Examiner

Tim T. Vo
SPE